

Functionalism, Qualia, and Other Minds

ELLIOT REED

How do we know that other people are conscious? This “problem of other minds” has traditionally been answered by citing others’ behavior, for the power of the intentional stance to predict and explain human behavior is unrivaled by any other system of explanation. I will argue, however, that the traditional solution succeeds only insofar as consciousness is conceived in terms of functional relations. If consciousness is conceived in terms of intrinsic, ineffable, indescribable qualitative states, the problem of other minds is unsolvable. Not only does the behavior of others fail to prove that they have qualia, it provides *no evidence whatsoever* for that contention, and neither does any other argument. Consequently, I will argue, we ought to conceive of consciousness in purely functional terms.

I will begin by stating some assumptions. First, I will assume that we are able to describe, predict, and explain, a great deal of others’ (and our own) behavior in terms of mentalistic categories: beliefs, desires, feelings, and so on. That is, Dennett’s assertion that taking the “intentional stance” yields predictive and explanatory power not available by any other known method (Dennett 1987) is correct.¹ Furthermore, I will assume that the categories of the intentional stance are good enough that a future theory of psychology will not abolish folk psychology, contrary to Paul Churchland’s thesis (Churchland 1981). (If it didn’t, ‘how do we know that other people are conscious?’ would be an empty question. One can only know things that are true.)

The first characteristic of the mental that I would like to consider is our ability to use the intentional stance. Now, if the intentional stance *works* in the deep way I am supposing it does, then we can treat folk psychology as a largely true (but incomplete) theory of the behavior of people.² The theory of folk psychology, unfortunately, does not provide us with exceptionless universal laws: at best we get rough-and-ready generalizations, but such lack of rigor is hardly a deathblow to the theory. Insofar as it is such a theory, it can be treated as a set of relations between observables (behaviors, etc.) and mental

states, so that being in a certain mental state provides dispositions to certain actions and to other mental states. In other words, mental categories are useful for prediction and explanation because we can treat them as functional categories.

OBJECTION: Why should we believe that mental categories are functional categories?

REPLY: Suppose we could not treat mental categories as functional categories. Then mental states would *not* provide dispositions to behaviors, or they would not provide dispositions to have other mental states. If this assertion were true, we would be unable to use them to predict others' behavior and mental states, which we are obviously able to do. Conversely, suppose that in addition to treating mental categories as functional categories, we need to consider some of their other properties in order to predict and explain the behavior of others using the intentional stance. But (by hypothesis) the functional characterization of mental states contains all the information how they cause dispositions to behavior, so no other quality of the state can be necessary in order to predict behavior. What of explanation? A characteristic of mental states that was useful for the explanation of observable phenomena, but that did not (*ceteris paribus*) explain that those phenomena were more likely than other phenomena would be no explanation at all! I conclude, therefore, that insofar as mental states are relevant to the prediction of observable phenomena such as behavior, they are relevant in virtue of their functional properties.

OBJECTION: Yes, the mental concepts of folk psychology do provide dispositions to behavior and other mental states, but why should we believe that they do this in a way that can be described as functionalist?

REPLY: Once we have established that the mental states of folk psychology provide dispositions to behave in certain ways and to have other mental states, we have established that folk psychology can be described functionally. The functional states of the theory are mental states, the outputs are behaviors, and the inputs are environmental stimuli.

Nonetheless, we cannot at present describe folk psychology as a functional theory. That is, we know how to apply folk psychological concepts, but not how to explain their

use (at least not very well). The development of such a description would require a substantial joint research program in linguistics and psychology. We would need research in order to determine the conditions under which people are willing to apply various mental predicates to other people, and further research to describe the sorts of predictive and explanatory inferences that people make on the basis of those applications.

By hypothesis, we are able to use the theory of folk psychology to predict and explain the behavior of others, but we cannot make the same inferences by using any other method. Consequently, we can safely conclude that other people are functionally equivalent to beings with mental properties. Unfortunately, this line of reasoning alone will not get us to the conclusion we want: namely, that other people *have* mental properties, since, in general, functional equivalence is not sufficient for equality. We need to introduce some other principle.

We get the cleanest argument by accepting the functionalist hypothesis. If being functionally equivalent to a conscious entity *is* being a conscious entity, then the remainder of our proof comprises one line: People are functionally equivalent to beings with minds; ergo, they have minds. Thus the problem of other minds is solved.

OBJECTION: What is 'functionally equivalent'?

REPLY: I will distinguish between two definitions of functional equivalence. The first definition of equivalence defines two systems as functionally equivalent iff both can be adequately described using a functional description where the same causal relations hold between the functional states and the inputs and outputs of the system and the description of both systems make reference to the same set of inputs and outputs. I will term systems that are equivalent in this sense 'I/O-equivalent.' Unfortunately, this definition of equivalence, when plugged into the functionalist thesis I give above, is apt to make functionalism parochial by denying mental states to entities that certainly have them.³ Consequently, I am led to accept another notion of equivalence in my definition. Two systems will be called A-equivalent iff both can be adequately described using a functional description and the description of both systems where the same causal relations hold between the functional states and

the inputs and outputs of the system, but the inputs and outputs need not be the same for both systems. Thus A-equivalency is necessary, but not sufficient, for I/O-equivalency. The same terms can be applied to functional states using analogous definitions. The functionalist thesis I am pushing, then, is that to be conscious is to be A-equivalent (or nearly A-equivalent) to a system that is well-described using folk psychology.

OBJECTION: If we can define having conscious states as being a certain type of functional system, then we will be able to find a functional characterization for *every* sufficiently complex system (the Atlantic Ocean, say) according to which it is conscious. And the fact of the matter is that the Atlantic Ocean is not conscious, no matter what functional characterization of its causal relations we attribute to it. Surely countenancing mental-state attributions of this kind is a *reductio ad absurdum* of functionalism!

REPLY: Those who make this objection are right to point out that, on the face of it, attributing mental states to a system such as the Atlantic Ocean seems bizarre. In fact, virtually every speaker of English (with the exception of a handful of animists, philosophers, and cognitive scientists) would agree the claim is definitely false. In the face of such overwhelming agreement (and given my complete inability to produce any arguments in favor of the alternative position), I will concede the point that, given the way words like 'conscious' are used in modern English, the claim I have been advancing is not true.⁴ Nonetheless, I maintain, it is still the best solution to the problem of other minds.

What an absurdity, to attempt to solve a philosophical problem by embracing an abjectly false claim! Nonetheless, I will argue, the problem we face here is not a problem with functionalism, but with our customary way of speaking. Accepting liberalism would require a substantial alteration in the way we speak about mental properties (as it would attribute such properties to a vast number of entities to which we do not now attribute them), but, I will argue, there are good reasons to accept such a change in our customary manner of speaking, and no such reasons (aside from inertia) to preserve it.

So what are the advantages of redefining mental predicates in purely functional terms? Taking the intentional stance allows us to obtain a substantial amount of predictive and

explanatory power that is not (presently) available to us by any other means. The redefinition I am proposing would not remove any of this power from the intentional stance, and if embracing it leads to the sort of research program I outlined earlier, it might well expand it. Treating beliefs, goals, desires, and so on as purely functional states would not prohibit us from using the intentional stance to predict and explain the actions of others just as we always have, even to the point of enabling us to understand (as best we can) others' comments about the taste of buttered toast.

Furthermore, as Dennett is wont to point out, the predictive power of the intentional stance is a matter of objective fact. It follows that if the hypothesis (that the functionalist thesis I am pushing leads to massive liberalism) is true, then at least some properties of many systems can be described using the intentional stance. In the case of most systems, this fact is probably a mere curiosity – if there are patterns in the Atlantic Ocean that can be described using the intentional stance, they are probably of no real interest to us – but it is possible that we will one day discover systems that can be described in a useful way using categories with functional properties very similar to those of the intentional stance, so that a wide variety of their input/output relations can be usefully described using the intentional stance. Such a possibility may sound fantastical, but the universe contains a great many complex systems. So a liberal ascription of conscious mental states may allow us to gain understanding, and we will lose no predictive or explanatory power that we already possess by adopting such a change of language.

OBJECTION: Consider the famous 'Absent Qualia Argument,' which has myriad variations, all of which essentially run like this:

- 1) The functionalist hypothesis is that having mental states merely requires a certain sort of functional state.
- 2) For any functional state, we can imagine a zombie that is in that functional state, but has no conscious, first-person, qualitative experience (i.e. has no qualia).
- 3) Having qualia is intrinsic to at least some mental states.
- 4) Therefore, a system need not be conscious merely in virtue of having certain functional states.

REPLY: The argument made by many (e.g. Dennett 1991) who

wish to deny (4) involves denying both (2) and (3) by denying that qualia exist. Such arguments tend to depend on examples designed to spur intuitions, and I will shamelessly admit that I find most of the literature on this subject confused. I will therefore ignore this particular counter-counterargument in favor of a more roundabout approach.

Consider that the question we are (ultimately) attempting to answer is 'how do we know that other people are conscious?' We have already shown that the functional portion of the folk psychological theory of mind is the only portion of the theory of mind that does any work in predicting or explaining any observable phenomenon. If qualia are important to our story about the mind, then, the problem of other minds reduces to the question, 'how do we know that other people really have qualia, rather than being unconscious zombies who fool us by virtue of being I/O-equivalent to conscious beings?'

And so we run into the problem: there is no way to distinguish between zombies and entities with qualia! There is not a shred of evidence (or any other kind) that I can point to to support the contention that George Bush has qualia, or to support the contention that he is a zombie. Consequently, if we demand that our notion of a mental state include having qualia, we are lead to the unfortunate conclusion that we *do not* know that other people are conscious!

OBJECTION: In effect, what you are saying is that the problem of other minds would be solved if we meant something else by 'minds', as your proposed recharacterization of consciousness removes the essential element - qualia! Two plus two would equal seven if we meant something different by 'two,' too, but you can't prove ' $2 + 2 = 7$ ' by redefining '2!'

REPLY: A functionalist redefinition of 'consciousness' will not be sufficient to solve the problem of other minds for those for whom the possession of consciousness requires ineffable, indescribable qualia. Since there is no way to demonstrate (or even provide evidence for) the proposition that other people have such qualia, it follows that the redefinition I am now proposing will not suffice as such a demonstration.

None of this need mean that talk of qualia (as distinguished from talk of 'qualia') need vanish. Under my proposed program of redefinition, we can continue to talk about

immediate sensory impressions, feelings, and every other state that is supposed to be a qualia. The difference is that in a liberal theory, we will make such ascriptions to any system for which it is functionally appropriate. If we make a robot that simulates a human being, and it eats buttered toast, we need have no qualms talking about how the toast tastes to the robot, just as we have no qualms about discussing how toast tastes to other people.

OBJECTION: You have shown that qualia are not necessary for any sort of scientific or everyday prediction or explanation, but there is more to life than prediction and explanation. That agents really have qualitative experience is important to many moral, aesthetic, and other types of value systems. Perhaps we can accept a language from which the idea of really having qualia has been expunged for the purpose of science, but it cannot be eliminated from value theory.

REPLY: I have already demonstrated that, using the conception of qualia employed by people who make these sorts of objections, there are no grounds for believing that other people are real experiencers with real qualitative states, rather than being mere zombies. Consequently, a value system that bases its evaluations on the possession of qualia makes it impossible to determine whose experiences are actually deserving of consideration and whose are not. Such a system cannot be used in making evaluations, and I see no reason to keep our old habits of speech simply because they allow us to preserve the illusion that certain unemployable systems of valuation are employable. Those to whom these sorts of considerations are compelling may, if they wish, continue to demand that having qualia be a necessary condition for consciousness in their idiolects, as long as they admit that their ascriptions of qualia to others rests on a leap of faith.

OBJECTION: John Searle offers an argument that suggests that other people *do* have qualitative mental states. I fear doing injustice to his argument by summarizing it, so I will simply quote him:

If you think for a moment about how we know that dogs and cats are conscious, and that computers and cars are not conscious . . . you will

see that the basis of our certainty is not "behavior," but rather a certain causal conception of how the world works. One can see that dogs and cats are in certain important respects relevantly similar to us. Those are eyes, this is skin, these are ears, etc. . . . behavior *by itself* is of no interest to us; it is rather *the combination of behavior with the knowledge of the causal underpinnings of the behavior* that form the basis of our knowledge. (Searle 1992 p.22)

So the thesis that other people are conscious is perhaps not proven, but its prior probability is much higher than that of the thesis that other people are zombies. I can reason from my own knowledge of my own qualia (we will suppose for a moment that I have such knowledge), and my picture of the world as consisting of particles in fields of force that my qualia are caused by the particular particles and fields of force in my brain. Thus, I can reason that other entities with similar brains have similar qualitative experiences.

REPLY: The problem with Searle's argument is that it rests on a seemingly dubious intuition about the sorts of brains that would produce qualitative mental states. As Dennett comments, "Perhaps left-handers [sic] brains, for instance, only mimic the control powers of brains that produce genuine Intentionality!"⁵ (Dennett 1987 p. 334) Undoubtedly Searle would find Dennett's supposition preposterous, but I doubt he could succeed in giving any reasons *why* it was preposterous without begging the question. He could not begin an empirical investigation to show that both right and left-handers have the brain characteristics that produce genuine qualitative experiences: in order to make such a determination, he would need some independent test to distinguish zombies from real conscious beings. Since no such test exists, Searle is flat out of luck.

OBJECTION: Your counter to Searle's response is just a crude form of anti-inductivist skepticism. We do not have any evidence that physical systems act as if there are electrons because electrons really exist, as opposed to the theory that physical systems act as they do because they are functionally equivalent to systems with electrons, but electrons are not real, but this does

not stop us from inferring (correctly) that electrons exist. Why hold qualia to a higher standard?

REPLY: In physics the properties of electrons are limited to the properties that functionally connect them to other elements of the theory of physics. A demonstration that a physical system is (perfectly) I/O-equivalent to a system with electrons is a demonstration that the system has properties with all the characteristics we ascribe to electrons. Qualia, by contrast, are supposed to have characteristics above and beyond their functional relations, and there can be no evidence that entities with all the properties of qualia (e.g. "ineffability"), rather than just their functional properties, exist.

OBJECTION: Block points out that there are many mentalistic categories whose conditions of application cannot be purely functional, because they require a certain type of relationship with the world. (Block 1978) For example, knowledge of some proposition *p* requires that *p* be true, and perception of some entity *E* requires that there actually be some *E* (if there is not, then the "perception" is a hallucination or mistake of some form). In general, propositional attitudes cannot be defined purely by A-equivalence, because having a propositional attitude requires some causal connection to the content expressed in the proposition. For example, in Hilary Putnam's Twin Earth example (Putnam 1973), the states of the Earthling who believes water is wet and the Twin Earthling who believes XYZ is wet are A-equivalent (or almost A-equivalent), but they are not both the belief that water is wet. One is causally connected to water; the other is causally connected to XYZ.

REPLY: Insofar as the application of mental predicates requires that some proposition not about the person be true, I am inclined to grant the objection, but it is no serious blow to the theory I am proposing to grant that some mental states are contingent on the facts of the world. We can always separate out the functionally relevant characteristics of the mental state (e.g. the belief that water is wet, as opposed to knowledge that water is wet) in order to predict and explain behavior, so there is no real problem.

The objection dealing with the connection of propositional attitudes to the propositions the attitudes are about is more significant. One way around the problem would be to accept that both the Earthling and the Twin Earthling *do* believe

the same thing, but that would be perverse at best: one of the reasons I gave for accepting this recharacterization of our intentional vocabulary was that it would allow us to continue using folk psychology as it applies to ordinary people without any alteration. The correct response, I believe, is to deny that functional states that are (almost) A-equivalent are necessarily the same mental state. So the beliefs of the Earthling and the Twin Earthling are A-equivalent, but they are not the same belief.

This point suggests that we need some sort of characterization of what the identity criteria for mental states are, since A-equivalence is insufficient. My proposal would be that the development/discovery of such criteria would be part of the job of the research program I outlined earlier. Finding criteria that both conformed to our intuitions about mental states (at least when applied to the case of other people) and were useful for science would be an important goal of such a program.

Luckily, our ability to apply the intentional stance need not wait for such a program to achieve success. We can describe the behavior of other people in intentional terms without it. Furthermore, we have a test for determining that entities we find in the world are conscious: if their behavior is well described in a wide variety of situations using folk psychology, then they are conscious. Since other people pass the test, they are conscious, and the problem of other minds is solved.

I have proposed a redefinition of the language we use to describe mental life. This language would redefine our mental terms so that the functional relations between them were preserved, but would concede that these terms can be applied to any system in which the same functional relations hold. All the predictive and explanatory power of the intentional language we use now would be preserved, although many important questions would be left to future researchers. The problem of other minds would be solved (since we know that other peoples' behavior is well-described using the categories of folk psychology) at the "price" of conceding that at least some things which most speakers of modern English would not call conscious are conscious. The notion of 'qualia' as "ineffable" or "indescribable" would vanish from the theory of mind, not because qualia have been shown not to exist, but simply because such talk lacks any predictive or explanatory value. Some

philosophers will complain that theories generated using this language will fail to explain why people have conscious experiences, and, given their notion of 'conscious experiences,' they will be correct. But an inability to explain a 'fact' that we can't evidence and don't need to predict or explain anything is of little concern to me.

Notes

¹ Undoubtedly many philosophers will think this assertion is far too weak and would prefer that I make some stronger claim. But I think very few (except perhaps the Churchlands) will be inclined to doubt it.

² I am not introducing the supposition that folk psychology is only a theory of the behavior of other people, but merely asserting that it is at least such a theory.

³ For an explication and defense of this point, see Block (1978).

⁴ Discussions with Benjamin George have convinced me that I am probably conceding too much by granting that functionalism is not true given modern English usage. But since my contention would only be strengthened if my thesis were true (or if its truth-value were ambiguous), I will grant the point anyway.

⁵ Dennett is talking about Intentionality, but the same point can be made about qualia.

Bibliography

- Block, Ned. "Troubles With Functionalism." In Savage, C.W., Ed., *Perception and Cognition. Issues in the Foundations of Psychology. Minnesota Studies in the Philosophy of Science*, vol.9. Minneapolis: University of Minnesota Press, 1978, 261-325. Rpt. Block, Ned, Ed., *Readings in Philosophy of Psychology*, vol. 1. 261-326. Cambridge, MA: Harvard University Press, 1980.
- Churchland, Paul. "Eliminative Materialism and the Propositional Attitudes." *The Journal of Philosophy* 78: 1981. pp. 67-90.
- Dennett, Daniel. *The Intentional Stance*. Cambridge, MA: MIT Press, 1987.
- Dennett, Daniel. *Consciousness Explained*. Boston: Little, Brown, and Company, 1991.
- Putnam, Hilary. "Meaning and Reference." *The Journal of Philosophy* 70: 1973. pp. 699-711.
- Searle, John. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press, 1992.